音楽情報科学 33－5
(1999. 12. 12)

隠れマルコフモデルによる音楽演奏からの音符列の推定

齋藤 直樹　中井 満　下平 博　嵯峨山 茂樹

北陸先端科学技術大学院大学 情報科学研究科

〒923-1292 石川県能美郡辰口町旭台 1-1

URL: http://www-ks.jaist.ac.jp/index-j.html

あらまし　本稿では、隠れマルコフモデルを用いて人間によって、鍵盤演奏された音符音長系列情報 (スタンダード MIDI ファイル) から意図された音符列を推定する手法を提案し、実験によりその有効性を実証する。人間が音楽演奏するときの各音符の物理的長さは、音符の正規の音長から意識的・無意識的に揺らぐため、楽譜投入・自動採譜などでは、意図された各音符の音価を正しく推定するのは容易ではない。本研究では、連続音声認識の定式化に倣って、演奏入力を音楽的に理解する原理を隠れマルコフモデル (HMM) によりモデル化し、意図された音符列を推定する。更に、同じ原理によりテンポ変化推定・小節線推定・拍子推定を提案する。評価実験により、一般に用いられている閾値処理より良好な結果が得られることを示す。

キーワード　●隠れマルコフモデル ●リズム認識 ●テンポ・小節線・拍子推定 ●自動採譜

Hidden Markov Model for Restoration of Musical Note Sequence from the Performance

Naoki Saitou　*Mitsuru Nakai*

Hiroshi Shimodaira　*Shigeki Sagayama*

Japan Advanced Institute of Science and Technology

1-1 Asahi-dai, Tatsu-no-kuchi, Ishikawa 923-1292

URL: http://www-ks.jaist.ac.jp/

Abstract　**This paper proposes the use of Hidden Markov Model (HMM) for restoration of a music note sequence from the music performance by human (reprenseted by a standard MIDI file). Successful experimental results are also presented. As the physical duration of a musical note in a human music performance fluctuates, intentionally or unintentionally, from the nominal length of the note, it is not easy to estimate the intended sequence of notes in autonmatic music transciption or music entry to computers. In the present paper, utilizing the formulation of continuous speech recognition, we use Hidden Markov Model (HMM) for modeling the process of the human understanding music performances and estimate the intended sequence of musical notes. We also apply the same principle to tempo estimation, bar line allocation, and beat estimation. Through experimental evaluation, we show the proposed method outperfoms existing methods.**

Key words　● hidden Markov model ● automatic rhythm recognition ● tempo, bars, and beat estimation ● automatic music transcription

1　まえがき

楽譜の浄書やMIDI演奏を目的にして、コンピュータへ楽譜を投入するソフトウェアツールが普及している。しかし、鍵盤入力から演奏者 (ユーザ) の意図した楽譜に変換するのは単純な問題ではない。たとえばMIDI鍵盤入力の場合、音高情報は正確に得られるが、音価 (音符の長さ) は (MIDIの時間分解能を単位として) ほぼ連続的な値として得られ、それを単純に処理しただけでは、意図された音符は得られない。その理由は、ユーザの演奏において、意図した音符の正規の長さから、実際に演奏した音符長には長短のずれを含むからである。よほどの熟達者

ですら、2分音符から16分音符までを機械的に正確に弾き分けるのは困難である。まして、音楽初心者が演奏する場合、テンポや正規の音符長に対し忠実に演奏することができない場合が多い。

音響信号入力からの自動採譜 [1] では、この問題はさらに困難になる。採譜システムとしては主にMIDI信号を対象とし音楽的分析を行うシステムと音響信号から周波数解析・音楽的分析を行い、様々な音楽解釈から楽譜を推定する手法がある [1]。これらは一般に人間の演奏情報を対象としている。楽譜化を目的とした演奏でない限り、曲のスタイル・表情付け、演奏者の音楽意図などにより、テンポや音長は意識的な変動を受ける。

以上のように、さまざまな音長変動要因が音長系列から音符シンボル列への変換を困難にする。従来手法や市販品の殆どは閾値処理をベースとしている。しかしそのような単純な処理では、ある市販ソフトウェアによる図1の例のように誤って音符推定される。演奏者の音楽的意図は同図左のようであり、同図右の物理的演奏情報に忠実な変換は必ずしも実用的ではない。この揺らぎに対して補正する研究は幾つか報告されており、閾値処理をベースとして、ヒストグラム処理による基準拍の設定手法、音楽的・文法的な強制或いはフレーズなどのルールの付加、またはテンポ情報を閾値設定に用いるものなどがある [2, 3, 4, 5]。また自動演奏という視点から、演奏情報と楽譜情報との比較から演奏の表情規則を抽出し、その規則により表情付けされた演奏からの採譜システムとして応用しているもの [6] や曲のビートを解析するビートトラッキングをマルチエージェントによりモデルベースで音楽的解析を行う報告もされている [7, 8]。

本稿では、同種の問題を扱っている連続音声認識分野の方法論をこの問題に適用し、音符列推定 (リズム認識)、演奏テンポ推定、拍子・拍節認識についてその定式化と実験結果について述べる。

2 HMMによる音符列推定

2.1 連続音声認識問題との同型性

本研究では、揺らぎのある音長列から音符列を推定する問題を、ボトムアップ的にずれを持つ音長をいかに音符に割り振るかを考える手法でなく、トップダウン的にどのような音符を意図して演奏した結果、入力演奏が観測されるかを仮説検証する、または解釈するという音声認識で成功している考え方を用いる。

そこで、整数関係にある正規の音符長が演奏によって揺らぎを持つ音長に変換される過程 (音長系列生成過程) を確率モデル化し、その逆問題として音符列を推定する問題を考える (図2)。具体的には、2レベルの確率モデルを作成し、それを基に HMM(Hidden Markov Model) を用いて音長系列生成モデルを作成する。HMM では尤度最大の原理によって音長系列が生成する遷移系列の中で最も尤度が高い系列をViterbi 探索によって求める。これによって、トップダウンアプローチで入力演奏を音楽的に解釈し、音長やテンポの揺らぎに頑健な推定を可能にする。

HMM は音声認識 [9, 10] において広く用いられているモデルで、本問題と連続音声認識は表1のように同種の問題と考えることができ、HMM を用いて尤度最大の状態遷移系列を探索 (Viterbi 経路探索) することにより音符列を求める問題として定式化できる。

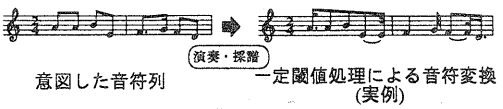

図 1: 閾値処理による音符への誤変換の例

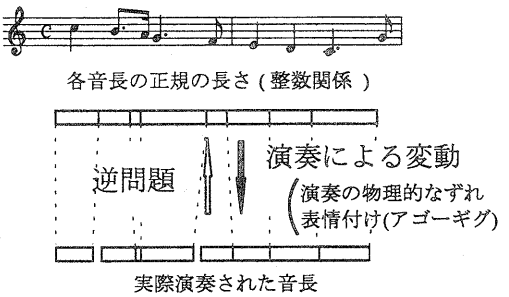

図 2: 逆問題としての音符列推定

表 1: 音声認識とリズム認識の対応

	連続音声認識	音楽リズム認識
入力単位	文音声	楽曲
語彙	単語	リズムパターン
隠れ状態	音響イベント	音符
観測値	スペクトル列	物理的音符長列

2.2 音符列モデル

音長に揺らぎがある演奏でも、聴き手には意図した音符列 (さらに、時には伸縮の意図も) が伝わるのはなぜか。これは聴き手は出現しうる音符列に関する常識を持っているからであろう。たとえば図1右のような楽譜は理論上は可能ではあるが常識に合わない。そこで、聴き手や音楽家の常識をモデル化するために、本手法では音楽的な制約として音符の推移をモデル化する。

これは音声認識における言語モデルあるいは文法に相当する部分である。ここでは簡単のため以下の2種類の音符列モデルを扱う。

- 2音符連鎖 (bigram) 確率モデル: 図3に示すように、任意の音符 i に任意の音符 j がそれぞれ確率 $a_{i,j}$ で後続するモデルである。制約力は弱いが、どんなリズムパターンにも対処できる。

- リズムパターンモデル: 図4に示すように、「リズム語彙」を定義し、リズムパターンの連鎖により曲が成立しているとするモデルであ

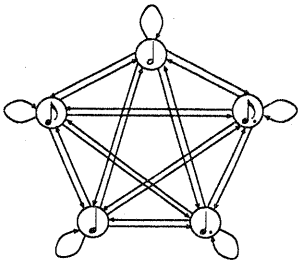

図 3: 音符連接のリズムモデル例

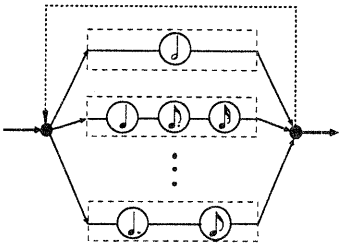

図 4: 2 拍単位パターンのリズムモデル例

る。このモデルは状態滞留確率を 0 とする点で、音声認識の HMM と若干異なる。

これらのモデルパラメータは、楽曲データから学習することができる。これは、人間の音楽経験による常識の形成に譬えられる。このようないわば「リズム文法」は、複雑に精度良く作成するほど、リズムパターン認識精度は向上する。また、これらはモデル楽曲のジャンルやスタイルに依存する。たとえば、ジャズのスィングリズムは、西洋古典派音楽として捉えると、演奏者が下手であると理解される。

実際に、童謡・民謡・歌曲 [11, 12, 13] を対象に 4/4 拍子の曲 88 曲より音符連接確率及びリズムパターンの統計を取った。パターンの分類として 1 小節単位パターンと 2 拍単位パターンの 2 種類を作成し、リズムパターンの種類は 1 小節単位パターン 267 種類、2 拍単位パターン 137 種類が得られた。また 3/4 拍子についても同様に 25 曲から統計をとり、1 小節単位パターン 68 種類が得られた。表 2 に例を示す。

2.3 音長の伸縮変動モデル

同一の音価の音符でも、既に述べたさまざまな要因により、その物理的音長が変動する。単純化して考えるため、これらを確率変動と見なそう。

図 5 に、テンポ指定つき演奏実験で得られた約 50 の演奏のデータから、4 分音符、8 分音符、符点 4 分音符の音長ヒストグラムの例を示す。横軸 (tick) は指定テンポにおいての 4 分音符の分解能を示す。今回は 4 分音符を 480 ticks として統計をとった。

本稿では、各音符の音長の分布を正規分布で近似する。正規分布の平均 μ は各音符長の正規の長さとし、標準偏差 σ は正規の音符長に比例する分と、固定分の和 $\sigma = a\mu + b$ の形で与えられると仮定する。a は、統計結果から、各音符の分散が音符が長い程広がるということに基づいた音符間での分散の相違を示し、b はどの音符でも人間の演奏内に含まれる

表 2: 音符列パターンの出現頻度例 (4/4 拍子)

頻度順	1 小節単位	2 拍単位
1 位	♩ ♩ ♩ ♩	♩ ♩
⋮		
10 位	♩. ♪♩ ♫	♪♩ ♪
⋮		

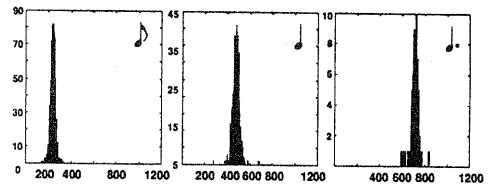

図 5: テンポ指定時の演奏の音長分布 (1/960 秒単位)

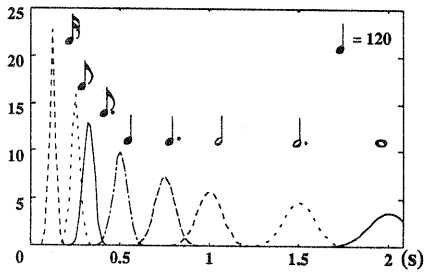

図 6: 各音符音長の変動モデル

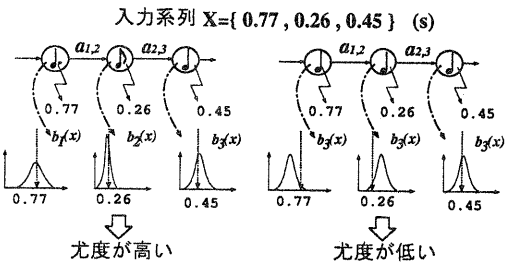

図 7: HMM による音符列推定の原理

固定分の物理的なずれを表す。図 5 から最小二乗法で得られた実験式は $\sigma = 0.05\mu + 0.011$ (秒単位) である。

しかし、実際の演奏テンポとモデルが仮定するテンポとにミスマッチがあること、この分布は演奏者に依存すること、また統計サンプル数が多くないことなどを考慮して、モデルの標準偏差を若干広めに設定し、以下では $\sigma = 0.06\mu + 0.0114$(秒) として以下でモデル化に用いている。これを図 6 に示す。このように音符 j が音長 x で演奏される正規分布確率密度を $b_j(x)$ と書く。確率モデルパラメータは、上のように演奏データから学習することができる。これは、人間の音楽経験による音長の揺らぎの常識の形成に譬えられる。

2.4 逆問題としての音符列推定

上記の2階層の確率モデルにより、意図した音符列 Q を演奏すると、音長時系列が X として観測される確率が求められる。すなわち、音長系列 X の生成確率 $P(X|Q)$ は上記の2つの確率の積で表すことができ

$$P(X|Q) = \prod_{t=1}^{N} a_{q_{t-1},q_t} \cdot b_{q_t}(x_t)$$

となる。q_t は時刻 t における音符の種類である。逆に、演奏情報 X が音符列 Q を意図したものである確率 $P(Q|X)$ は、Bayes の定理

$$P(Q|X) = \frac{P(X|Q)P(Q)}{P(X)}$$

によって、$P(X|Q)P(Q)$ を求める問題と考えることができ、先の音長系列生成確率を求めることになる。ここで $P(Q)$ は音符列 Q が生成される確率であるので、$P(Q)$ をリズムパターンの連結確率としてモデルに組み込む。

3 HMMによる音符列推定

3.1 HMMによる音符列推定

$P(X|Q)P(Q)$ を求めるため HMM を用いて、2つの確率モデルを統合し、最も尤もらしい音符列を推定することができる。HMM によるモデル化において各パラメータは以下のような意味を持つ。

- 状態 s_i：音符 i
- 初期確率 π_i：ある音符 i から曲が始まる確率
- 遷移確率 $a_{i,j}$：音符 i から音符 j へ遷移する確率
- 出力確率 $b_j(x)$：音符 j が音長 x で演奏される確率
- 入力系列 X: 演奏された音符長系列 $X = \{x_1, \cdots, x_n\}$

図7に HMM による音符列推定の概念図を示す。演奏された音長系列 $X = \{0.77, 0.26, 0.45\}$(秒) が入力された時、この系列 X を生成する確率が最も大きい音符列 Q を、隠れ状態系列の Viterbi 探索により求めることができる。

最小単位となるリズムモデルを作成する。図3と4のような音符連接モデルとリズムパターンモデルの2つを作成し、HMM の出力確率を音長の変動モデルに相当させる。リズム推定においては最小単位となるリズムの連結として楽譜を推定する。

3.2 HMMによる音符列推定実験

モデル: 4/4 拍子の曲から統計をとった音符連接モデル及び2拍単位リズムパターンモデル (図 3, 4) を用いた。得られたパターン数は 2.2 節で述べた通りである。

入力: 楽譜投入の際に演奏者がテンポ通りに演奏できないことを想定した実験条件として、

条件 1：テンポ指定ありでなるべく忠実な演奏

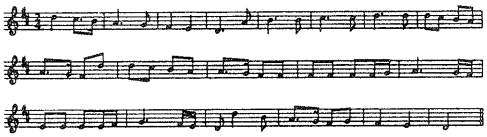

図 8: 入力曲「もろびとこぞりて」[12]

表 3: 音符列推定精度: 20 曲の音符認識率 (%)

method	休符挿入	休符削除
閾値処理 (XGworks)	40.70	(85.86)
閾値処理 (Finale)	38.79	(87.20)
音符連接 HMM	53.73	87.39
2拍 Rhythm HMM	59.65	97.26

条件 2：テンポ指定なしでテンポ一定の演奏
条件 3：テンポ指定なしでテンポ変動を含む演奏

について扱った。

条件 1 の演奏について被験者 10 名 (合計 16 曲) に対し、音符列推定実験を行った。実験の対象曲としては、よく知られていて比較的短く音符の種類が豊富な「もろびとこぞりて (ニ長調)」(図 8)[12] を選んだ。

評価方法: 実験では MIDI 信号を対象としているので、実験の評価としては、各音長が正しく音符変換されているかのみを評価する。正解精度は以下により評価する。

$$accuracy = \frac{N - sub - del - ins}{N} \times 100(\%)$$

- N：未知入力の総音符数
- sub：誤った音符に置換された数
- del：正しい音符が脱落した誤り数
- ins：異なる音符が挿入された誤り数

実験結果: 音符列推定結果を表 3 に示す。同音反復の場合などに演奏に短いポーズが挿入されるが、そのまま音符列推定を行った場合 (表中「休符挿入」) と、閾値処理により除いて処理した場合 (表中「休符削除」) の両方の場合についての認識率を示す。これらの短いポーズを放置すると、評価上では挿入誤りが増加して認識率が低下する。

閾値処理 (XGworks および Finale) の場合は不要なタイや短い休符が多く出現するが、実験の主旨により、タイで表現されている部分についての記譜誤りは除き、また、不要な 16 分休符を除いて集計した場合を表中の「休符削除」の欄に掲載した。

4 HMMによるテンポ推定

4.1 固定テンポ/変動テンポ推定問題

上記のリズムパターンモデルは、時間情報として各音符音長がとりうる値を出力確率に対応させたモデル化であるため、ある一定のテンポの入力のみ解析可能である。そこで、各リズムパターンモデルを

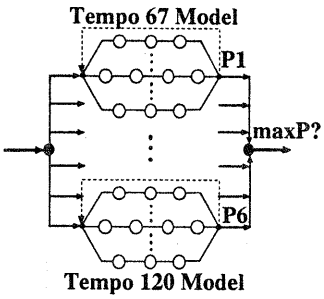

図 9: 一定テンポモデル

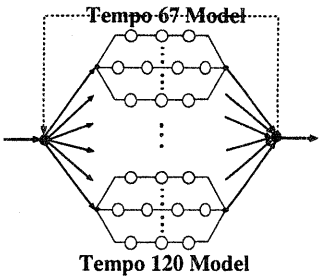

図 10: 変動テンポモデル

複数のテンポ毎に作成し、入力に対して各テンポ毎に並列に尤度計算を行い、尤度が最大となるテンポを推定結果とすることでテンポによる適用範囲を広げる (図 9)。テンポは 67〜120 の間で対数的に 5 分割し、6 つのテンポを採用した。

テンポの前後の揺らぎが激しい入力に対処するために、図 10 のように、図 9 の一定テンポモデル間に遷移確率を設け、階層型 HMM を作成する。これにより、移り変わるテンポに追従した解析を可能にする。

4.2 固定テンポ推定実験

入力: 入力は条件 2(テンポ指定なしでテンポ一定の演奏)「もろびとこぞりて」[12] について被験者 10 人 (10 演奏) を対象とする。用いるモデルは図 9 の一定テンポモデルにより、6 つの固定テンポ候補中から演奏されたテンポを一つ推定する。

評価方法: 演奏は奏者の演奏技術による揺らぎ以外の表情付けなどの変動要因は含まないことをふまえ、その曲全体が演奏された平均テンポ (1 分間の四分音符の数) を

$$\text{演奏テンポ} = \text{拍数}/\text{演奏時間 (分)}$$

により定義し、比較対象とする。

テンポ推定結果: 曲の演奏時間から求めた平均テンポと一定テンポ HMM の選択されたモデル (最も尤度が高いモデル) を表 4 に示す。認識率によって多少異なる場合もあるが、6 種のテンポのうち一番近いモデルが選択され、テンポ推定率は 100% であった。

表 4: テンポ推定結果 (条件 2: 10 曲。A: 拍数 (38 個)/演奏時間 (分), B: 一定テンポ HMM)

player#	A	B	player#	A	B
1	98.35	95	6	116.41	120
2	93.31	95	7	111.74	107
3	99.20	95	8	99.88	95
4	127.06	120	9	109.25	107
5	106.34	107	10	65.16	67

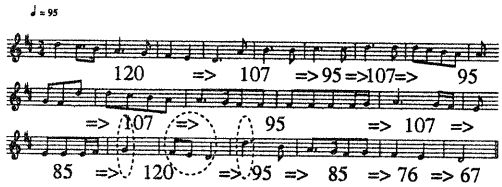

図 11: 変動するテンポと音符列推定 (○は誤推定)

4.3 テンポ変動認識実験

入力: 同じ入力曲 (条件 3) で大幅なテンポ変動を含む演奏に対する実験を行う。モデルは図 10 に示す変動テンポモデルを用いる。一番多く採用されたテンポのモデルをその曲が演奏された平均のテンポとする。

テンポ変動問題に対する推定結果: 図 11 に、意図的に極端なテンポ変動を行った演奏に対するテンポ変動推定実験結果を示す。尤度最大の状態遷移系列をたどると、以下のテンポモデル間の遷移を行っていることがわかった。

Tempo 120 (初期モデル) → 120 → 120
→ 107 → 107 → 95 → 107 → 95 → 95
→ 107 → 95 → 95 → 107 → 85 → 120
→ 120 → 95 →85 → 76 → 67

極端に遅い演奏個所では、音価は倍にテンポは速めに推定された結果、誤推定が生じたが、妥当な推定であるとも考えられる。2 拍単位パターンモデルなので小節毎にテンポが推移するような場合は、小節内での急激な変化や、小節毎に誤推定されたりすることがある。テンポ間の遷移確率を調整することにより、この誤認識が減少できる可能性がある。

5 HMM による拍節推定

5.1 拍子/開始拍/小節線位置推定問題

演奏から楽譜を復元する場合には、音符列のみならず拍子の推定、開始拍 (アウフタクトかどうか) の推定、すなわち小節線をどのように入れればよいかという問題を解決する必要がある。これらの問題も、以上に述べた確率モデルによって定式化できる。

拍子特性が顕著に現れるのは、1 小節中に含まれる音符パターンであると考える。そこで 4/4 拍子、3/4 拍子毎に 1 小節 1 パターンのリズム統計をとり、各モデルで入力された旋律の尤度を並列計算し音符列を推定する。ここで尤度最大の原理を利用し、尤

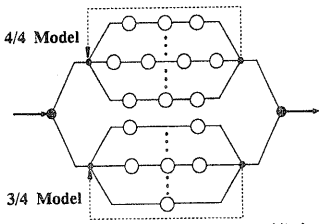

図 12: モデルによる拍子推定

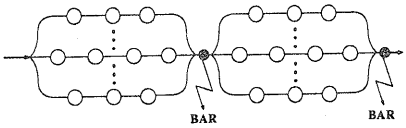

図 13: モデルによる小節線推定

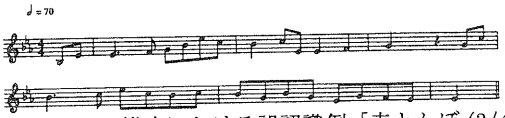

図 14: 拍子推定における誤認識例「赤とんぼ (3/4 拍子)」— リズムパターンの観点からは妥当な解

度が高い遷移系列を求めその系列が 4/4 であるか 3/4 であるかを判定し、拍子推定結果とする。

小節線推定は、事後処理による挿入とモデルを用いた挿入方法の 2 種類を試みた。事後処理による挿入では、拍子情報を基に曲の冒頭から拍数分カウントし挿入する。モデルによる推定手法では、図 13 のように 1 小節 1 パターンのリズムモデルを用いる。これにより、各リズムパターンの最終状態が選択された後、小節線を挿入する。

アウフタクト (上げ拍) の可能性も含めた小節線位置の推定では、事後処理による推定では、最後に数があわない場合にアウフタクト (上げ拍) であると判断し最初にもどり 2 つ目の音符からカウントを始める。モデルでは、アウフタクトの小節を初期確率のみ持つ別のリズムパターンとして与える。これにより、曲の途中でそのパターンが選択されることを防げる。

5.2 拍節推定実験

モデルと入力データ: 図 12 のモデルを用い、4/4 拍子 10 曲、3/4 拍 10 曲に対し条件 1 の演奏を入力した。リズムの最小単位としては双方とも 1 小節単位パターンのモデルを用いた。2 拍 1 パターンのものはパターン 2 つにつき小節線を出力という形式で行った。

拍子・小節線推定結果: 4/4 拍子については 10 曲全てについて正しく拍子推定できた。3/4 拍子 10 曲中 8 曲は正しく推定できたが、残る 2 曲は音符列としては正しく推定されたが、拍子は 4/4 拍子と誤推定された。図 14 に誤推定例を示す。リズムパターンとしては、1 フレーズが 3 小節になっているところに違和感があるが、4/4 拍子と考えても矛盾はない。このような場合の拍子推定は、旋律あるいは想定される和声まで含めたさらに高度な総合モデルが必要となる。

本手法の小節線推定では、拍子を誤推定した場合、小節線は本来の楽譜と全く違う箇所に挿入される。また、拍子推定が正しくとも、音符列 (リズムパターン) が正しいかどうかによって小節線位置の推定結果も変わる。

6 まとめと今後の課題

本稿では、音楽演奏の音符音長系列データに対し、連続音声認識の方法論を適用して統合的な確率モデルと最尤経路探索により、意図された音符リズム推定、テンポ推定、拍子推定、小節線位置推定などが統一的に行えることを示した。

今後は、ジャンルやスタイルを考慮 (に依存) したリズムパターンのモデル学習方法、楽曲フレーズのようなより大きな曲構造を反映したモデル、未知リズムパターンへの対処 (音声認識における未知語対策に対応)、リズムパターンに依存した音長伸縮特性を考慮した推定 (同じく文脈依存モデルに対応)、ユーザのスキルや癖を学習するユーザ適応技術 (同じく話者適応に対応)、A*アルゴリズムなどの効率的な解探索、N-best アルゴリズムの適用などの発展により本法の適用可能性を広げたい。さらに、音響信号入力に対して適用し、自動採譜の一要素技術として用いたい。

参考文献

[1] 長嶋洋一, 橘本周司, 平賀護, 平田圭二: コンピュータと音楽の世界, bit 別冊, 共立出版株式会社, 1998.

[2] H. C. ronguet-Higgins: Mental Processes, The Mit Press, 1987.

[3] 片寄, 井口: "知的採譜システム," 人工知能学会誌, Vol.5, No.1, pp.59-66, 1990.

[4] 海野, 中西: "音楽情景分析における楽音認識と自動採譜," インタラクション 99 予行集, 1999.

[5] P. Desain, H. Honing: "Quantization of Musical Time; A Connectionist Approach," Computer Music Journal, Vol. 13, pp. 56–66, 1989.

[6] 野池, 乾, 野瀬, 小谷: "演奏情報と楽譜情報の対からの演奏表情規則の獲得とその応用," 情報処理学会音楽情報科学研究会, 97-MUS-26-16, pp.109-114, 1998.

[7] 後藤真孝, 村岡洋一: "音楽音響信号を対象としたビートトラッキングシステム -小節線の検出と打楽器音の有無に応じた音楽的知識の選択-," 情報処理学会音楽情報研究会, 97-MUS-21-8, pp.45-52, 1997.

[8] Masataka Goto and Yoichi Muraoka: "Real-time Rhythm Tracking for Drumless Audio Signals – Chord Change Detection for Musical Decisions –," IJCAL-97 Workshop on Computational Auditory Scene Analysis, 1997.

[9] 中川聖一: 確率モデルによる音声認識, 電子情報通信学会, 1988.

[10] L. Rabiner, B.-H. Juang: Fundamentals of Speech Recognition, Prentice-Hall, 1993.

[11] 中学生の音楽 1,2,3, 教育芸術社, 1983-85.

[12] 楽しく歌おう, 神奈川県中学校音楽教育研究会, 1983.

[13] 世界名歌 110 曲集, 全音楽譜出版社.

[14] YAMAHA XGworks V 3.0, ヤマハ株式会社.

[15] Finale 3.5.2r2 for Windows, Coda Music Technology, 1987–1995.